R E P O R T R E S U M E S

ED 020 546 CG 002 127

A REVIEW OF RESEARCH ON THE WELSH FIGURE PREFERENCE TEST.

BY- AIKEN, LEWIS R., JR.
RICHARDSON FOUNDATION, GREENSBORO, N.C.

PUB DATE JUN 67

EDRS PRICE MF-$0.25 HC-$1.36 32P.

DESCRIPTORS- *PERSONALITY STUDIES, *PERSONALITY TESTS, PERSONALITY, CREATIVITY RESEARCH, CONFORMITY, *OBJECTIVE TESTS, MATURATION, WELSH FIGURE PREFERENCE TEST, THE REVISED ART SCALE, MINNESOTA MULTIPHASIC PERSONALITY INVEN, CALIFORNIA PSYCHOLOGICAL INVENTORY, BARRON WELSH ART SCALE

THE WELSH FIGURE PREFERENCE TEST (WFPT) CONSISTS OF 400 BLACK AND WHITE LINE DRAWINGS. THE EXAMINEE INDICATED WHETHER HE LIKES OR DISLIKES EACH DRAWING. THIS NON-LANGUAGE, OBJECTIVE METHOD OF ASSESSING PERSONALITY REVEALS TWO MAJOR FACTORS--(1) THE TENDENCY TO LIKE OR DISLIKE FIGURES, AND (2) THE TENDENCY TO PREFER SIMPLE, SYMMETRICAL FIGURES. STUDIES DEMONSTRATE THAT THOSE WHO PREFER SIMPLE DESIGNS TEND TO BE ANTISOCIAL OR REBELLIOUS. RESPONSES ON THE WFPT ARE DISCUSSED IN RELATIONSHIP TO MENTAL MATURITY, CONFORMANCE, INDEPENDENCE OF JUDGMENT, MASCULINITY-FEMININITY, PERSONALITY INVENTORIES, AND THE DEVIATION HYPOTHESIS. REVIEWS OF THE WFPT INDICATE THAT ALTHOUGH THE TEST ITEMS ARE RELATIVELY CONTENT-FREE AND DIFFERENTIATE BETWEEN GROUPS, THEY DO NOT PREDICT WELL IN INDIVIDUAL CASES. WFPT SCORES REPORTEDLY DO NOT HAVE HIGH CORRELATIONS WITH RATINGS OF ACTUAL BEHAVIOR. THE BARRON-WELSH ART SCALE AND THE REVISED ART SCALE ARE BRIEFLY REVIEWED. (PH)

ED020546

A REVIEW OF RESEARCH ON THE WELSH FIGURE PREFERENCE TEST

By

Lewis R. Aiken, Jr.

Guilford College

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Published By

The Creativity Research Institute

Of The

Richardson Foundation, Inc.

June, 1967

CG 002 127

Greensboro, N.C.

A REVIEW OF RESEARCH ON THE WELSH FIGURE PREFERENCE TEST[1]

Lewis R. Aiken, Jr.

Guilford College

The Welsh Figure Preference Test, published by Consulting Psychologists Press, consists of 400 black and white line drawings of varying complexity. In its original form, the drawings on the test were presented individually on cards, but the subsequent booklet form, consisting of eight drawings per page, has been used more extensively. The examinee is instructed to indicate whether he likes or doesn't like each of the drawings, either by marking an "L" or a "D" on the separate answer sheet, a task which takes about 40 to 50 minutes. The Preliminary Manual for the test (Welsh, 1959) states that it can be used with persons of age six and above and as many as 27 different scores obtained: DL (Don't Like Scale), RP (Repeat Scale), CF (Conformance Scale), BW (Barron-Welsh Art Scale), RA (Revised Art Scale), MF (Male-Female Scale), NP (Neuropsychiatric Scale), CN (Children Scale), in addition to 12 a priori content scores, an MV (Movement Scale) score, an FG (Figure-Ground Scale) score, and five sex symbol scores.

Welsh's purpose in constructing the test was to provide a non-language, objective method of assessing personality. In a sense, the Welsh Figure Preference Test (WFPT) was to be a non-language form of the MMPI, a test which had proven so successful with literates. Initially (Welsh, 1949), 143 people, six of whom were artists, sorted several hundred line drawings, drawn in black ink on 3 by 5 cards, into two piles—those that they liked and those that they disliked. After these sorts were obtained, a factor analysis of the preferences revealed two major factors: (I) the tendency to like or dislike figures (acceptance-rejection factor), and (II) the tendency to prefer simple, symmetrical figures as opposed to complex, asymmetrical ones. The second bipolar factor was orthogonal to the first.

Further study of the responses to these designs (Welsh, 1949) indicated that individuals who preferred the simple, symmetrical figures tended to be conservative and conventional, whereas those who preferred the complex, asymmetrical designs tended to be more antisocial or rebellious. Psychologists' ratings of these two groups of individuals depicted the first group as more enthusiastic, optimistic, conservative, organized, and conventional, whereas those in the

[1]The writing of this review was supported by the Creativity Research Institute of the Richardson Foundation.

second group were described as more cynical, pessimistic, depreciative, overtly hostile, and socially dissident.

Reviews and Criticisms

The Sixth Mental Measurements Yearbook (Buros, 1965) contains two reviews of the WFPT. The review by Gordon Anderson of the earlier card form of the test appeared first in the Journal of Consulting Psychology (1960) and is concerned only with some of the initial research on the test. Anderson refers to the test's ultimate intent of providing an index of emotional adjustment and identifying and quantifying personality characteristics. He then proceeds to describe the various scales, concluding that the test is primarily a research tool in need of further validation work.

In his review of the booklet form, Harold Borko makes a number of criticisms of the WFPT. After summarizing reliability coefficients obtained for various scales—.94 and .90 for the RA Scale, .64 and .74 for the MV Scale, and .92 to .51 for the DL Scale—Borko concludes that adequate reliability studies have not been done. In addition, it is alleged that the standardizing populations were too small—100 male VA patients, 75 normal adult males, 75 normal adult females, and 42 boys and 40 girls aged six to eight. Although Borko refers to Welsh's (1959) statement that the Barron-Welsh Art Scale is in final form, it is also noted that the Revised Art Scale is not in final form, the Male-Female Scale has not held up under cross-validation, the Children Scale has not been validated, and the Neuropsychiatric Scale has not been cross-validated. Borko concludes that "the test does not measure many of the personality directions," and even those that are measured by the test "... are not measured with much confidence" (p. 198). He recommends further empirical research to determine which sets of items differentiate among various groups, since he feels that the test presently "... has little practical value" (p. 198).

Both reviews of the WFPT indicate that although the test items are relatively content-free and do differentiate between groups, they don't predict well in the individual case. In addition, the WFPT scores reportedly do not have high correlations with ratings of actual behavior.

The major portion of the present review will be concerned with a consideration of the scales on which the greatest amount of research has been done. We shall not discuss all of the 27 scales listed in the 1959 manual, since the majority of these have received little research attention. We shall not be limited, however, to discussing only a sample of these 27 scales, because several new scales have been devised since the manual was published. Finally,

studies concerned with the relationship of WFPT scores to other tests and research which has used the WFPT to test Berg's "deviation hypothesis" will be summarized briefly.

Barron-Welsh Art Scale

It is proper to consider the Barron-Welsh Art Scale (BW Art Scale) first, because more research has been conducted with this scale and the Revised Art Scale than all other scales combined. The original form of the BW Art Scale consisted of 40 items disliked by artists significantly more often than people in general and 25 items liked by artists significantly more often. The 40 items which the artists disliked were of the simple-symmetrical type, and the majority of the 25 items which they liked were of the complex-asymmetrical sort. It was found that total score on this 65-item scale differentiated between artists and non-artists, a finding which was cross-validated on two other groups (Barron & Welsh, 1952).

Reliability of the BW Art Scale

Wrightsman (1964) investigated the test-retest reliability of the BW Art Scale. For a group of 75 college-student examinees who took the test twice over an average interval of 166 days, the reliability coefficient for total score on the BW Art Scale was .799. Furthermore, the reliability of the "Don't Like" items on the scale was .761, and the reliability of the "Like" items was .782.

Relationships of BW Art Scale to Personality Characteristics

In a discussion of the earlier work with the BW Art Scale, Barron (1953a) summarized the correlates of the scale with the Adjective Check List and an attitude scale. One significant finding was a positive correlation, for a group of male graduate students, between rated originality and BW Art Scale scores. It is reported that preference for complexity (high scores) on the BW Art Scale was positively related to rapid personal tempo, verbal fluency, impulsiveness, expansiveness, independence of judgment, originality, and breadth of interest, but negatively related to rigidity, control of impulse by repression, social conformity, ethnocentrism, and political-economic conservatism.

In another experiment, using an Asch-type arrangement, Barron (1953b) found that low scorers on the BW Art Scale tended to "yield" to group pressure by giving perceptually inaccurate reports, whereas high scorers tended to be more "independent" in their judgments, giving perceptually accurate reports in spite of pressure from the group to do otherwise.

In his book, Creativity and Psychological Health (1963), Barron again summarized research on the BW Art Scale. The following are some of the investigations referred to in that book. One study (Barron, 1952) was concerned with the personality characteristics of 40 male graduate students, some of whom had scored high and others low on the BW Art Scale. Both groups described themselves on Gough's Adjective Check List, with low scorers on the BW Art Scale marking the adjectives "contented, gentle, conservative, patient, peaceable, serious, individualistic, stable, worrying, thrifty, dreamy, deliberate, moderate, modest, responsible, foresighted, and conscientious" more often, and high scorers on the BW Art Scale marking the adjectives "gloomy, pessimistic, bitter, dissatisfied, emotional, pleasure-seeking, unstable, cool, irritable, aloof, sarcastic, spendthrift, distractible, demanding, indifferent, anxious, opinionated, temperamental, and quick" more often. Ratings obtained from others revealed that high scorers on the BW Art Scale were perceived as being more original and artistically expressive and as having better aesthetic judgment than low scorers. The responses of the subjects to an attitude scale showed low scorers on the BW Art Scale to be more socially conforming, more respectful toward tradition, more patriotic, and more morally judgmental than high scorers on the art scale.

In a related investigation, Schultz and Knapp (1959) administered a 62-card form of the BW Art Scale and a 100-item self-description word list to 367 male, non-psychiatric VA patients. The subjects were divided into four groups according to their scores on the BW Art Scale: (I) Like both symmetrical and asymmetrical designs, (II) Like symmetrical and dislike asymmetrical designs, (III) Dislike symmetrical and like asymmetrical designs, and (IV) Dislike both symmetrical and asymmetrical designs. Then these four groups were compared with respect to the adjectives which they checked most frequently as descriptive of themselves. The results were analyzed by comparing the responses of group II to those of groups I, III, and IV combined, and the responses of group III to those of groups I, II, and IV combined. Group II, who preferred symmetrical to asymmetrical designs, more often described themselves as "always looking on the dark side, always on the go, impatient, on edge, play it safe, push myself, self-conscious, talker, tired, and worn out." Group III, who preferred asymmetrical to symmetrical designs, more often described themselves as "broke, different than others, forgetful, like to figure things out, nervous, never had the breaks, pretty clear about what I believe, rugged, talker, and thinker."

We shall refer briefly to three other investigations concerned with the relationships between BW Art Scale scores and personality. Martin (1954) used a 40-item form of the BW Art Scale and correlated the scores with a measure of

intolerance of ambiguity. The latter measure consisted of counting the number of questions that the subjects asked in attempting to clarify an ambiguous situation. An essentially zero correlation was found between BW Art Scale scores and the latter measure.

Secrest and Jackson (1961) administered a forced-choice form of the BW Art Scale to 60 first year nursing students by randomly pairing simple and complex figures. The corrected split-half reliability of the test in this format was .92. A correlation of .35, significant at the .05 level, was found between scores on this form of the art scale and reputed "social intelligence" for these subjects. Finally, Pine (1962), in a study employing 12 male students, found a correlation of .74 between scores on the BW Art Scale and a measure of "adaptive regression", and a correlation of .64 between art scale scores and a measure entitled "control of primary process".

Preferences for Paintings

Barron (1952) obtained the preferences of 40 graduate students for 105 postcard-size reproductions of paintings by famous European artists and related these preferences to scores on the BW Art Scale. It was found that low scorers on the BW Art Scale, who liked symmetrical designs, preferred paintings depicting high-born and holy people—paintings concerned with religion, authority, and aristocracy. This group disliked paintings of women of low birth and occupation and radically experimental, unnatural, or esoteric paintings. The reverse was true of students who made high scores on the BW Art Scale, and consequently preferred asymmetrical designs. This second group approved of the experimental, modern, primitive, and sensual paintings, and disliked religious, traditional, aristocratic, and emotionally controlled paintings.

Another investigation along these same lines (Child, 1962) was concerned primarily with the relationship between the ratings of the aesthetic values of paintings by 22 undergraduate males and expert judges' ratings of the aesthetic values of the paintings. Although the students and the experts' judgments showed very little relationship, there was a correlation of .45 between BW Art Scale scores and the preferences of the students for paintings, scored according to a scale of aesthetic value established from the experts' judgments.

Studies of Artists, Architects, and Scientists

Rosen (1955) administered the BW Art Scale to 44 art students, eight art faculty members, and a group of non-artists matched with the artists on age and sex. The mean

BW Scale score of the advanced art students was 40.7, the mean of the beginning art students 39.0, the mean of the art faculty 41.1, and the mean of the non-artists 22.1. Thus, there were only small differences between the means of the art students and art faculty but a large difference between the means of the non-artist and artist groups. Another finding was that course grades of the art students correlated .34 with BW Art Scale scores, and ratings of the students' originality correlated .40 with art scale scores.

Some of the results of MacKinnon's (1961a) study of architects and research scientists are presented in Table 1. From this table it may be seen that the mean BW Art Scale score of 40 "creative" architects was 37.1, whereas the means of the two less creative architect groups were 29.5 and 26.1. The mean of 15 "creative" scientists was 30.7, and the means for the less creative scientist groups were 22.1 and 19.2. Thus, mean scores on the BW Art Scale differentiated between subsamples of architects and research scientists differing in level of creativity. MacKinnon (1961a) reports a correlation of .48 between rated creativity and scores on the BW Art Scale for these groups.

Studies of Other Occupational Groups

Additional support for the hypothesis of a positive relationship between creativity and scores on the BW Art Scale was obtained by Barron (1955, 1957) in an assessment program which employed 100 Air Force captains as subjects. Eight tests, or measures, of originality (Unusual Uses, Consequences, Plot Titles, Rorschach O+, TAT: Originality, Anagrams, Word Synthesis Originality, and Inkblot Originality) were administered to each subject, and the sum of the standard scores on these tests was related to several other variables. The investigation focused on two groups—those scoring high on the eight-test composite (15 Ss, "Original" group) and those scoring low on the eight-test composite (15 Ss, "Unoriginal" group). On the BW Art Scale, the mean of the Originals was 19.40 and the mean of the Unoriginals 12.67, a significant difference. In a further analysis of the data, however, Barron (1957) found no positive correlation between the BW Art Scale and the eight-test originality composite when scores on Terman's Concept Mastery Test were partialled out.

MacKinnon (1961b) summarizes data relating BW Art Scale scores to creativity in the occupational groups referred to above and others, and some of these findings are also presented in Table 1. For example, the mean BW Art Scale score of a sample of 16 "creative" women mathematicians was 28.1, and the mean of 28 "representative" women mathematicians 26.9. Highly-creative industrial researchers scored a mean of 30.7,

Table 1

Mean Scores on the Barron-Welsh Art Scale for Various Groups

Sample	Number of Subjects	Mean	Standard Deviation
Artists (standardization group)[a]	80	40.3	12.9
Artists (first cross-validation group)[a]	30	39.1	13.8
Architects I ("creative" architects)	40	37.1	9.8
Writers ("creative")	19	32.9	11.1
Team members of first American expedition to attempt Mount Everest	15	31.5	12.1
Research scientists I ("creative" scientists"	15	30.7	6.3
Architects II (control sample for number 3 above)	43	29.5	10.1
Women mathematicians I ("creative")	16	28.1	12.5
Women mathematicians II ("representative")	28	26.9	15.4
Men mathematicians I ("creative")	26	26.9	12.7
Architects III ("representative" American architects"	41	26.1	12.1
Research scientists II ("less creative")	15	22.1	14.1
Men mathematicians ("less creative")	21	19.4	10.1
Scientists III ("least creative")	15	19.2	8.7
Unselected adult males	343	13.9	11.2

Note.- After Barron, 1963b, p. 3.

[a]The corrected means on the revised, 62-item form of the BW Art Scale are slightly lower than these.

whereas the mean of industrial researchers of lower creativity was only 19.2. It is interesting to note that the mean score on the BW Art Scale of a group of student engineers was only 21.5, a fact which MacKinnon interprets as indicating that student engineers need stimulation of their aesthetic interests if they are to become creative.

Finally, Gough (1961) obtained a correlation of .41 between scores on the BW Art Scale and ratings of creativity in 45 physical science and engineering researchers.[2]

Relationships to Other Tests

In this section we shall complete our review of correlational studies on the BW Art Scale by summarizing briefly the results of two investigations. Bieri, Bradburn, and Galinsky (1958) administered the BW Art Scale and the Embedded Figures Test to groups of 62 female and 50 male undergraduates. In the case of the males there was a significant negative correlation ($r = -.36$, $p<.01$) between speed on the Embedded Figures Test and BW Art Scale scores. The corresponding correlation for women, however, was only -.08. The investigators interpreted this result as indicating that men who are more proficient in coping with the complex figures of the Embedded Figures Test prefer more complexity in art productions.

Caracena and King (1962), in an exploration of the unitary nature of complexity as a personality dimension, obtained a score for each of 30 male and 30 female undergraduates on the BW Art Scale, seven scores from the Bieri-Blacker scoring modification of the Rorschach, and a score on Kelly's Repertory Test. The correlations of the other eight measures of complexity with BW Art Scale scores ranged from .12 to -.12, all of which were insignificant. Apparently, complexity is not a unitary dimension of personality.

The Teaching of Creativity

Until now we have dealt exclusively with correlational studies concerning the validity of the BW Art Scale. Recently, however, Brown (1964, 1965, 1966) has reported a series of experiments in which changes in BW Art Scale scores were used as the dependent, or criterion, variable. In the first of these experiments (Brown, 1964), the teacher in a semester

---

[2]Recently, Professor Wayne Holtzman has communicated with the writer concerning the results of a study on the ability of the WFPT to discriminate among a select group of artists, architects, and engineers. Using a sample of 28 artists, 31 architects, and 26 engineers, significant differences were obtained among the mean scores of the three groups on six of the a priori content keys and the Movement Scale of the WFPT. Of the three groups, the mean of the architects

course for elementary education juniors attempted to develop a creativity "self" or "frame of reference", built around a series of "creativity symbols", in the students. The purpose was to train the students so that the use of one of the symbols would "trigger" the creativity frame of reference. This teaching procedure resulted in a mean gain of eight to nine points on the BW Art Scale.

In a second study (Brown, 1965) a similar procedure was used to develop "creative sub-selves" which could be triggered by different symbols. The mean gain in BW Art Scale scores from initial testing in February to retesting in May was 18 points for the group when the creativity sub-selves were triggered at retesting time. The mean gain was significantly higher ($p<.001$) than gains obtained under conventional conditions.

Brown (1966) also reports on a project designed to produce changes in teachers' feelings about children and teaching, their colleagues, and their own potential for growth. Significant changes ($p<.005$) in BW Art Scale scores for 29 teachers in four schools were associated with this project during its first year.

Criticism of the BW Art Scale

It is probably unnecessary to be too critical of studies that have employed the BW Art Scale, because many of the shortcomings of research with this scale are generally true of research on psychological tests. From the preceding review, however, it may be noted that the correlations of the BW Scale with other tests, although frequently significant, are not large, and the results are infrequently cross-validated. In addition, studies that have found differences between the means of various groups on the BW Art Scale are usually based on rather small samples. A fair conclusion after reviewing this work is that the results are interesting and provocative, but it is not clear at this point exactly what aspect or correlate of "creativity" is being measured by the BW Art Scale.[3]

Another criticism of the BW Art Scale was made by Moyles, Tuddenham, and Block (1965). These writers report

---

was highest and the artists lowest on the Freehand Line-Simple, Freehand Line-Complex, and Black scales, while the mean of the architects was highest and the engineers lowest on the Ruled Line-Simple, Ruled Line-Complex, and Shading scales. Finally, the mean score on the Movement Scale was highest for the engineers and lowest for the artists.

[3]See McWhinnie (1965) for a review of research on aesthetic measurement and perceptual choice and an indication of some of the contradictions in investigations on this subject.

that the stimulus dimension of simplicity-complexity, which is what Barron and others have frequently stated to be the dimension being measured by the BW Art Scale, is highly confounded with the dimension of symmetry-asymmetry. A separation of these two dimensions showed that they were, independently, of small importance in determining figure preference. As a consequence of their findings, the writers issued a word of caution against the tendency to attribute figure preference to the single dimension of simplicity-complexity and, by extension, to the simplicity and complexity of the personalities of examinees.

The Revised Art Scale

Welsh (1959) points to a difficulty with the BW Art Scale, stemming from the fact that the scoring standards on the 65-item scale are such that higher scores may be obtained when the examinee places more items in the Don't Like (DL) category than in the Like (L) category. For example, a subject can obtain a score of 40, a rather high score, simply by indicating that he doesn't like any of the items. Consequently, high correlations are obtained between DL score and total score on the BW Art Scale. In order to eliminate this difficulty, a Revised Art Scale (RA Scale) was constructed. The RA Scale consists of 60 items, 30 of which high scorers, as opposed to low scorers, on the BW Art Scale indicated that they like and 30 of which they indicated that they don't like. In constructing the scale, high and low groups of scorers who also had average DL scores were used. Thus, a scale was produced that had a high correlation with BW Art Scale scores but zero correlation with DL total scores.

Correlations of the RA Scale With Other Tests

It will be recalled from findings reported in an earlier section (Barron, 1957) that the BW Art Scale and Terman's Concept Mastery Test apparently have substantial overlap. The question arises, therefore, whether the RA Scale and tests of intelligence also have a great deal in common. If the answer is in the affirmative, then the effectiveness of the RA Scale as an independent measure of creativity would be somewhat reduced. A partial answer to the question was obtained by Welsh in a study of 368 gifted high school students in a residential summer school. For this group, the correlation between two measures of intelligence, one verbal (Terman Concept Mastery Test) and the other non-verbal (D-48), was significant. But the correlation of both tests with the RA Scale was essentially zero. It was also found that students in the Academic Division of the program scored higher in tested intelligence, but students in the Art Division made higher scores on the RA Scale. It should be noted, however, that these statistics were obtained on a highly select group and that the correlation between intelligence

and RA Scale scores, for example, might prove to be somewhat higher with groups showing a greater range of scores (see Welsh, 1966).

Millman and Chang (1966) correlated scores on the RA Scale with the scores of 32 undergraduates on the Graves Design Judgment Test and the Meier-Seashore Art Judgment Test. The rank-order correlation between the RA Scale and the Graves was .46, whereas the correlation between the RA Scale and the Meier was -.16, and between the Meier and Graves it was .19. The writers attribute the higher correlation between the RA Scale and the Graves as being due to the fact that both of these tests use abstract designs.

In a comparison of three creativity measures—the RA Scale, Mednick's Remote Associates Test, and a Similarities Test—Colman (1966) found a correlation of .52 ($p<.10$) between RA Scale scores and Remote Associates scores in a group of 13 students with high RA Scale scores. On the other hand, for a group of 11 students with low RA Scale scores, there was a correlation of .53 ($p<.10$) between RA Scale scores and mean number of words used per similarity on the Similarities Test.

Relationship of RA Scale to Creativity Motivation

Golann (1962a, 1962b) administered a "creativity motivation" questionnaire, which measured preference for creative activities, and the RA Scale to sixth and eighth grade children. Then the correlations between these two variables were computed for different age and sex groups. The obtained correlations varied from -.14 to +.37, four of them being significant at the .05 level or better. Perhaps these results could have been interpreted more meaningfully if Golann had also obtained measures of intelligence and/or school grades on his subjects. In any event, he also reported that males who made high RA scores preferred self-expressive, independent activities which allow for use of creative capacity, while males who made low RA scores preferred activities which were structured, assigned, familiar, or routine.

RA Scores and Artists

Raychaudhuri has issued a number of reports (Raychaudhuri, 1961, 1962, 1963, 1966a, 1966b, 1966c) on the RA scores of musical and painting artists, as compared to those of controls. In one investigation (Raychaudhuri, 1966a), the mean RA Scale score of 30 musical artists of India was 29.00, and the mean of 50 non-artists was 19.00, a significant difference ($p<.01$). The mean RA score of 60 painting artists, however, was 40.83, which is significantly higher ($p<.01$) than that of the musical artists. These obtained differences between artists and non-artists on the RA Scale are reinforced by other studies (Raychaudhuri, 1961, 1962, 1963, 1966c).

Raychaudhuri interprets the difference between the RA scores of the painters and musicians with the hypothesis that "... those who succeed in painting arts have greater ambiguity tolerance in visual field than that of those who succeed in arts employing a different sense-media, e.g. auditory in case of music" (sic, Raychaudhuri, 1966a, p. 3).

In a more extensive study of the personality factors differentiating 30 professional musicians with high scores on the RA Scale from 30 non-creative subjects with low scores on the RA Scale, Raychaudhuri (1966b) had two clinicians analyze the differences between the responses of these two groups to the Rorschach, the TAT, the Szondi Test, and a life-history interview. The results indicated that "The creative musician is more distinctly marked by his emotional and temperamental characteristics than by cognitive and motivational aspects of his personality structure" (Raychaudhuri, 1966b, p. 210). Thus, the musicians appeared to be more emotional, egocentric, exhibitionistic, and sensitive individuals who were stimulated by frustration and preferred activities and situations that permit a greater degree of individualism and self-expression. It was also concluded that familial and socio-cultural field factors were more important than heredity in determining musicianship.

Personality Correlates of RA Scale Scores

Cashdan and Welsh (1966) classified 311 adolescents attending the Governor's School of North Carolina into high ($RA \geq 46$) and low ($RA \leq 16$) creativity on the basis of their RA Scale scores and sex (Male/Female) and then analyzed the differences among these four groups in terms of their scores on the 24 scales of Gough's Adjective Check List (ACL). The results were that the two high RA groups made significantly higher scores on the lability, heterosexuality, change, autonomy, and counseling readiness scales of the ACL than the two low RA groups, while the two low RA groups scored significantly higher on the deference and order scales of the ACL. Regarding differences between the sexes, males scoring high on the RA Scale also scored significantly higher than high RA females on the self-confidence, dominance, exhibition, and autonomy scales of the ACL, while high RA females scored higher on the ACL nurturance, heterosexuality, succorance, abasement, deference, and counseling readiness scales. However, there were also significant RA by Sex interactions on the achievement, dominance, endurance, order, succorance, and abasement scales of the ACL.

In a second part of the Cashdan and Welsh (1966) study, all subjects who scored 40 or higher on the RA Scale were classified according to specialty (art versus natural science) and sex. The results revealed no significant differences in ACL scores between the two specialty or two sex groups.

Significant specialty by sex interactions occurred, however, on the favorability and defensiveness scales of the ACL. In a summary of these results, the writers characterized the highly creative adolescent as an independent, non-conforming, change-seeking individual who has open and active interpersonal relationships. In contrast, the adolescent of low creativity, and the male in particular, is described as a somewhat compulsive, highly achievement-motivated individual who seems to be experiencing external pressures to succeed. It is also concluded that creative adolescents of both sexes are quite similar in their personality characteristics.

Use of the RA Scale in Other Educational Settings

In a study of creative teachers and pupils, James (1963) obtained a mean RA score of 39.33 for the former groups, which is quite close to that obtained with artists. James also observed a greater tolerance of distorted and diverse figures on the part of these creative teachers.

A before-and-after study by Falcone (1962) used the AC Test of Creative Ability and the Welsh Figure Preference Test as both pretests and posttests. The purpose of the investigation was to evaluate changes in scores on these tests associated with a unit in a high school economics course designed to encourage students to develop their creative abilities. Materials used in the unit were a series of colored slides employed by the teachers in their regular classes. The findings were that the posttest mean (49.92) on the RA Scale of the WFPT was not significantly different from the pretest mean (50.00). In addition, there was a very low relationship between increases in T-scores on the RA Scale for the 77 students and whether or not they had prior experience in art or interior design. Finally, it should also be noted that the posttest mean (50.20) on the Conformance Scale of the WFPT was not significantly different from the pretest mean (50.01).

RA Scale Scores of Various Age Groups

Table 2 gives means and standard deviations of scores on the RA Scale as a function of age or grade level. Observe that, for both boys and girls, the means tend to increase slightly from the third grade through the seventh grade and then decline slightly through the tenth grade. Although the relatively high positive correlation (ρ = .62 for Golann's data) between means and standard deviations for boys should be taken into account in the statistical evaluation of the significance of the differences among means, there is at least a suggestion of a curvilinear trend of mean RA score across age group. If this trend were supported by other research it would certainly provide an opportunity for speculation. For example, if RA Scale scores are a measure of creative potential, can it be concluded that creative poten-

Table 2

Means and Standard Deviations for Various Age Groups on the Revised Art Scale of the Welsh Figure Preference Test

Sample	Number of Subjects	Mean	Standard Deviation
Four-to-six year olds[a]	20	26.30	5.43
Six-to-eight year olds[b]	82	22.59	9.10
School Children[c]			
Third grade boys	21	21.67	4.73
Third grade girls	32	21.81	12.62
Fourth grade boys	62	24.32	9.93
Fourth grade girls	79	23.29	10.02
Fifth grade boys	41	23.02	11.89
Fifth grade girls	45	24.09	12.30
Sixth grade boys	88	24.11	12.03
Sixth grade girls	99	24.38	12.55
Seventh grade boys	52	29.44	12.66
Seventh grade girls	68	28.13	12.60
Eighth grade boys	77	25.66	13.60
Eighth grade girls	124	26.83	12.78
Ninth grade boys	111	23.20	12.82
Ninth grade girls	96	23.30	12.76
Tenth grade boys	97	22.26	11.47
Tenth grade girls	86	21.52	13.82
Adults[b]			
Men	75	17.33	11.46
Women	75	19.33	12.20

[a]After Smith, 1962.

[b]After Welsh, 1959.

[c]After Golann, 1961.

tial reaches a maximum in junior high school and then declines? If this is true, what aspects of maturation or educational structure are responsible for the rise and fall of creative potential? Whatever the answers to these questions, the relationship of age to preference for designs on the WFPT should be a topic worthy of investigation. And it has been investigated, as will be seen in the next section.

WFPT Scores as a Function of Perceptual and Mental Maturity

The Children Scale

It was stated earlier in this review that the WFPT could be administered to children as young as six years of age. But since the task on the WFPT is relatively simple, it would seem that the test might even be administered to preschool children. And as Welsh (1959) indicates, the WFPT has been given to children as young as 3½ years. Smith (1962) notes, however, that children younger than 4½ years are usually unable to make the aesthetic discriminations required by the test. The preferences of preschool children are usually less reliable and distinctive than those of older children, because the former respond more immediately and affectively. Smith's (1962) study of 20 preschool children found that the Don't Like (DL) and RA Scale mean scores were actually higher than those of adults. But the investigator hypothesized that the distinctiveness of children's preferences for either simple, symmetrical figures or complex, asymmetrical ones increases with maturity.

The Preliminary Manual of the WFPT (Welsh, 1959) refers to a Children Scale of 50 items obtained by comparing the responses of 82 boys and girls, aged six to eight, with those of 150 people-in-general. Welsh notes, however, that the scale is only in preliminary form and has not been cross-validated.

A Perceptual Maturity Scale

Van de Castle (1962) attempted to confirm Hesterly and Berg's (1958) finding that the perceptual choices of children and adult schizophrenics are similar but that both groups differ from normal adults in their choices. The procedure was first to construct a "Perceptual Maturity Scale" consisting of 94 items from a 144-item short form of the WFPT which significantly differentiated between 200 elementary school pupils (second and third grades) and 200 college students having low Schizophrenic scores on the MMPI. Next, this "Perceptual Maturity Scale " (PMS) was administered to new groups of elementary school children, college students, and adult schizophrenics. The mean score on the PMS was signifi-

cantly higher for 11-12 year olds than for 7-8 year olds and higher for the college students than for the children. In addition, the mean score of the college students was significantly higher than that of the schizophrenics. But unlike one of the findings of Hesterly and Berg (1958) with the Perceptual Reaction Test (PRT), the mean score of the children on the PMS was also significantly lower than the mean of the schizophrenics. Van de Castle (1962) confirmed his results with a revised form of the PMS designed to eliminate the effects of "Don't Like" response bias that existed in the earlier form of the scale. He also concluded that Berg's PRT measures response bias rather than perceptual maturity.

Further work by Van de Castle (1965) and Edwards (1961) was oriented toward the development of a more comprehensive scale of perceptual maturity that would distinguish among different age groups. Edwards' (1961) procedure was to administer 150 items from the WFPT to 200 children aged eight years or less and to 200 normal adults in the age range 21-49 years. The 150 items were presented in a forced-choice format, each item-pair consisting of a probable adult choice and a child choice, based on previous work. The 92 item-pairs which significantly differentiated between the children and the adults were then assembled as a Perceptual Maturity Scale, and this scale was cross-validated on 100 eight to twelve year old children, 100 normal adults and 100 schizophrenic adults. It had been hypothesized that schizophrenic adults would make significantly lower scores than normal adults on such a scale, but this hypothesis was not confirmed. Therefore, the 72 out of 92 items which significantly differentiated between the children and the normal adults were reassembled as a new scale, and each item was given a one-point weight. The means of the total scores on this new scale showed a progressive increase with age, and they were significantly different for normal adults and children. Consistent with the results of the initial study, viz. the finding of an insignificant difference between the mean scores of normal and schizophrenic adults, however, was the finding of a significant difference between the mean scores of children and schizophrenic adults. In addition, maturity ratings of the children by their teachers were positively related, to a slight degree, to the PMS scores of the former, but there was no relationship between PMS scores and scores on tests of intelligence or reading ability.

In a further, more carefully controlled effort along these same lines, Van de Castle (1965) developed a 72-item, forced-choice form of Perceptual Maturity Scale. Scores on this form of the PMS had a reliability of .90 but zero correlation with IQ. Not only did mean scores on the scale show a steady increase from age seven through adulthood, but students rated as behaviorally more immature by their teachers had significantly lower mean PMS scores than students rated

as behaviorally more mature. Some other findings were:

(a) The mean PMS score of a group of delinquent males was significantly lower than that of a group of non-delinquent males of the same age.

(b) The mean PMS score of a group of adult schizophrenics was significantly lower than that of a group of normal adults.

(c) A group of students who were low on a "maturity scale" consisting of items from the California Psychological Inventory had a lower mean PMS score than a group of students who made high scores on the "maturity scale".

(d) For 42 college students, the correlation between PMS scores and scores on the Welsh Anxiety Scale of the MMPI was -.40.

Table 3 shows some of the results obtained in a later study with various age groups of Americans and Cuna Indians. It will be noted that, for both Americans and Cuna Indians, mean PMS score is an increasing function of age, although the mean scores of the Indians are lower than those of the Americans at every age.

These studies demonstrate that the PMS has a substantial degree of construct validity, although all of the investigations which have been reviewed are of the cross-sectional or correlational type. Van de Castle (1965) also notes that significant differences on the PMS appear only when the number of subjects is large and that differences are obtained between the scores of males and females at the younger age levels.

Mental Maturity and the WFPT

In an extensive, well-conducted study, Watson (1964)[4] took a slightly different approach to the relationship between maturity and responses on the WFPT. The subjects were 66 boys and 66 girls in each of three school class groups: gifted ($IQ \geq 125$ on the WISC), regular classes ($90 \leq IQ \leq 110$ on the California Test of Mental Maturity), and mentally retarded ($IQ \leq 75$) on the Stanford-Binet). The WFPT was administered to these 396 children, and chi-square comparisons were made between the responses of the three groups to the 400 items on the test, as well as their scores on 32 scales of the WFPT. As a result of the item analysis, two cross-validated scales were con-

[4] Watson (1964) also gives a good summary of the literature on the WFPT to 1964.

Table 3

Average Scores on the Perceptual Maturity Scale (PMS) of Americans and Cuna Indians at Various Age Levels

Age Level	Americans			Cuna Indians		
	No. of Males	No. of Females	Average PMS	No. of Males	No. of Females	Average PMS
10	85	85	45.11	11	2	38.45
11	85	85		27	27	
12	17	17	48.38	38	21	45.62
13	17	17		51	8	
14	19	19	54.13	53	33	50.38
15	19	19		58	16	
16	19	19	55.78	26	10	53.67
17	19	19		25	6	
18+ (Adult)	90	90	58.80	12	7	54.70
Total	370	370	50.76	301	130	47.93

Note.- Personal communication from Dr. Robert Van de Castle.

structed: (I) a scale of 103 items which discriminated mentally retarded children of low estimated mental age from the other groups, and (II) a scale of 50 items which discriminated gifted students of high estimated mental age from others. An analysis of the mean scores of these groups on 32 scales of the WFPT indicated that the mentally retarded made significantly different scores from others on the Conformance, Revised Art, Neuropsychiatric, Children, Male-Female, Movement, Shading, Dotted, Sex Symbol, and Mixed scales. In the main, the scores of the mentally retarded subjects suggested less conforming, more childlike responses, more consistent with the responses of the mentally ill than with those of normals. They disliked shaded figures, figures suggesting movement, and figures generally preferred by females. In contrast, the gifted subjects scored high on the Conformance and Revised Art scales and low on the Children and Neuropsychiatric scales. The answers of the gifted were more consistent with those of females in general, but they disliked the dotted and sex symbols, neutral figures. Another interesting finding was that, in general, the preferences of children in the regular classes were more consistent with those of the gifted than with those of the retardates. Finally, there was, as expected, a significant difference between the means of boys and girls on the Male-Female Scale, a difference which increased with age.

In sum, Watson (1964) demonstrated that the three mental ability groups responded differently to the items of the WFPT. Therefore, it might be reasonable to suppose that certain items on the WFPT could be used as a non-verbal intelligence scale. Again, however, the fact that the WFPT can discriminate between groups of different mental abilities should not be interpreted as meaning that it can differentiate between two specific individuals having different abilities. And Watson's (1964) conclusion that the WFPT can now be used in special education classes appears, for this reason, to be questionnable.

Don't Like, Conformance, and Male-Female Scales

Don't Like Scale

An examinee's Don't Like (DL) score on the WFPT is obtained simply by counting the number of items, out of 400, that he places in the "Don't Like" category. The Preliminary Manual of the WFPT (Welsh, 1959) reviews a series of unpublished investigations in which scores on the DL Scale were correlated with scores on Gough's Adjective Check List, the California Psychological Inventory, the Edwards Personal Preference Schedule, and the MMPI. The manual also summarizes the results of the study by Schultz and Knapp (1959) referred to earlier in this review. The reader will recall that this

study found that subjects who liked the symmetrical and disliked the asymmetrical designs on the BW Art Scale described themselves quite differently than subjects who liked the asymmetrical and disliked the symmetrical designs.

In a more recent study, concerned with the diagnosis of brain damage, L'Abate, Boelling, Hutton, and Mathews (1962) administered the WFPT and three other tests to 13 pairs of men and 16 pairs of women patients. Each pair of subjects consisted of a brain-damaged patient and a schizophrenic patient matched on sex, race, education, and length of hospitalization. The mean DL score for male schizophrenics was 144.00, as opposed to 117.92 for the brain-damaged men, and the mean DL score for female schizophrenics was 111.00, as opposed to 123.25 for the brain-damaged women. The differences between mean DL scores of the brain-damaged and schizophrenic groups were statistically insignificant, as were the differences between the means of the brain-damaged and schizophrenic patients on 26 other scales derived from the main score. Although the mean DL scores of both patient groups were quite a bit smaller than the means for normals reported by Welsh (1959), the scores did not appear to vary with age or length of hospitalization. L'Abate et al. (1962) concluded that DL Scale scores have questionnable usefulness as indices of brain damage.

Conformance and Independence of Judgment

The Conformance (CF) Scale, also referred to as "Conformity Scale" or "Agreement with Consensus Scale", of the WFPT consists of 38 items, 13 of which were liked and 25 disliked significantly more often by 50 artists than by a group of 150 people-in-general. Welsh (1959) reports a number of significant correlations between CF scores and scores on the Adjective Check List, the California Psychological Inventory, and the MMPI.

Lim (Unpublished paper, no date) views the CF Scale as a measure of consensus in perception, similar to the Rorschach popular response, which presumably measures the congruence between the examinee's perception and that of normal people. However, this measure of "perceptual consensus" did not differentiate between the Chinese and non-Chinese in a group of 140 children in a San Francisco school. Of course, these results may be interpreted as indicating either that the Chinese and non-Chinese children perceive the world in similar ways or that the CF Scale is not sensitive to differences in perception between the two groups.

Three of the variables used by Wrightsman and Cook (1965) in a factor analysis of 73 personality, attitude, and aptitude variables were Number of "Like" Responses, Agreement with Consensus, and Independence of Judgment on the WFPT. The subjects were 177 female college students in Nashville, Tenn-

essee. Eleven factors were extracted, and the following factor loadings were obtained for the three WFPT variables. Agreement with Consensus had a loading of .3527 and Independence of Judgment a loading of -.5865 on Factor I: Rigidity; Independence of Judgment had a loading of -.2527 on Factor III: Anti-Negro Attitudes; Agreement with Consensus had a loading of .3417 and Number of "Like" Responses a loading of -.4183 on Factor VI: Sociability.

Masculinity-Femininity and the WFPT

Welsh (1959) describes a Male-Female (MF) Scale on the WFPT which was constructed by contrasting the responses of 75 men with those of 75 women. Although the original 46-item MF Scale did not hold up under cross-validation, items from the WFPT were used subsequently in a number of investigations of the masculinity-femininity dimension. For example, Lessler (1964) used figures from the WFPT, along with other figures, in a study of the referents used in sorting line drawings into male and female categories. The results showed that some of the drawings (symbols) were sorted according to a cultural referent, others according to a genital referent, and still others according to both cultural and genital referents simultaneously. In general, however, the subjects responded to the symbols on the highest possible psychic level, i.e. in terms of cultural, rather than genital, referents when both were present simultaneously.

LaCrosse (1965) used stimuli from the MF Scale of the WFPT, along with others, in a study designed to determine the relationship of biological sex, selected psychological sex variables, intelligence, and perceptual rigidity to visual exploratory behavior in a sample of 20 girls and 20 boys in the second and sixth grades. The MF Scale designs were those preferred significantly more often by girls or boys and were selected to determine whether the complexity or sexual pull of a design would be the more important variable in determining total time spent looking at it. No relationship was found between intelligence, grade, or perceptual rigidity and visual exploratory behavior, viz. total time spent looking at the stimuli. In addition, there was no relationship between biological sex and total looking time, and the relationship between psychological measures of masculinity-femininity and time spent looking at the stimuli was complex. The findings of this investigation underscored the multidimensionality of the concept of masculinity-femininity and suggested further work designed to isolate its components.

Littlejohn (1966) investigated the responses of 324 ninth-grade boys and 332 ninth-grade girls to conventional and subtle measures of masculinity-femininity. The conventional measures were the MMPI Mf Scale, the California Psychological Inventory (CPI) Fe Scale, and Nichols' "Obvious"

Scale. The subtle scales were the Female-Male (FM) Scale of the WFPT—a 76-item, cross-validated version of the earlier Male-Female Scale—and Nichols' "Subtle" Scale. High scores on all of these scales are in the more feminine direction. The major hypothesis was that highly creative males and females —high creativity being defined as a score falling in the top third of the distribution of Revised Art Scale (RA) scores for the group of 656 subjects—would show a reversal of sex identification which would be manifested by their scores on the masculinity-femininity measures. A comparison group of low creative students consisted of those scoring in the bottom third of the distribution of RA scores. Concerning scores on the FM Scale, the results verified the hypothesis in the case of males: The 113 high-creative boys scored significantly higher (mean FM = 25.76). However, the difference between the means of the two male groups was not significant on any of the other masculinity-femininity measures. In the case of the girls, the hypothesis was not confirmed. In contrast to what was predicted, the 98 high-creative girls also scored significantly higher than the 107 low-creative girls on the FM Scale. The mean FM score for the former group was 57.63, but for the latter group it was only 36.64. Interestingly enough, however, the direction of the differences between the means of the two female groups on three of the other four measures of masculinity-femininity was just the reverse of that on the FM Scale, viz. in the direction hypothesized. Thus, low-creative girls scored significantly higher than high-creative girls on the "Subtle", "Obvious", and MMPI Mf scales.

Some related findings of Littlejohn's (1966) study are pertinent and of interest. Further comparisons between the means of high-creative and low-creative boys and girls on 11 other WFPT scales revealed that the means of the high-creatives was significantly higher than that of the low-creatives on the Barron-Welsh Art Scale, Children Scale, Movement Scale, Figure-Ground Scale, Sex Symbols-Male Scale, Sex Symbols-Female Scale, and Sex Symbols-Neutral Scale. On the other hand, the mean of the high creatives was significantly lower than that of the low-creatives on the Conformance Scale, Neuropsychiatric Scale, Sex Symbols-Combined Scale, and Sex Symbols-Mixed Scale. The directions of the differences between the means of high-creatives and low-creatives were the same for males and females.

The consistent differences between the means of high-creative and low-creative groups on various WFPT scales would indicate that the intercorrelations among the WFPT scales must be substantial. In fact, Littlejohn reports a correlation of .88 for the 324 males and .81 for the 332 females between RA and FM. Also, Harris (1961) gives intercorrelation matrices for 16 WFPT variables, and inspection of these matrices shows that the high degree of correlation among many

of the WFPT scales cannot be explained solely by overlap among items, because scales other than the BW Art Scale and RA Scale have few items in common.

Relationships of the WFPT to Personality Inventories

In this section we shall consider a number of correlates of responses to items and additional scales on the WFPT. Twice during this review we have referred to the Schultz and Knapp (1959) study, in which it was found that the adjectival self-descriptions given by subjects who preferred simple designs were different from those given by subjects who preferred complex designs. In a related investigation, Rosenberg and Zimet (1957) correlated scores on the California Public Opinion Scale (F Scale) with various scores on the WFPT for three small groups of students. It was found, as hypothesized, that high scorers on the F Scale, interpreted as being more "authoritarian" in outlook, preferred balanced, two-dimensional, commonplace designs. In contrast, low scorers on the F Scale, interpreted as being less "authoritarian" in outlook, preferred the more complex or deviant designs on the WFPT.

WFPT and the MMPI

It will be recalled that Welsh's (1959) expressed purpose in developing the WFPT was to construct a kind of non-verbal form of the MMPI which would be relatively free from faking and useful in personality assessment. Therefore, investigations which have related WFPT scores to MMPI scores are of particular interest. The major question here is: To what extent do WFPT scores measure the same variables that are measured by the MMPI?

Pepper (1957) investigated the relationship of the A ("anxiety") and R ("repression") scales of the MMPI to WFPT scales. Four groups of college students were tested: 9 students scoring high on both the MMPI A and R scales (Hi-Hi), 10 students scoring high on A but low on R (Hi-Lo), 20 students scoring low on A but high on R (Lo-Hi), and 6 students scoring low on both A and R (Lo-Lo). The results showed significant differences between the Lo-Hi group and the Hi-Lo group on the empirically derived Neuropsychiatric and BW Art scales and on the a priori content Ruled Line-Simple (angular, bilaterally symmetrical), Ruled Line-Total, Freehand Line-Complex, Simple-Total, and Shading and Cross-Hatched Line scales. The means of the Lo-Hi group were significantly lower than those of the Hi-Hi group on the Ruled Line-Simple, Ruled Line-Total, Simple-Total, and Black scales of the WFPT. In addition, there were significant differences between the Hi-Lo and Hi-Hi groups on the Neuropsychiatric, BW Art, Ruled Line-Simple, Freehand Line-Complex, Shading and Cross-Hatched Line, and Movement scales of the WFPT. Pepper (1957) concluded

that the figure preferences of the Lo-Hi group, who preferred simple figures and rejected complex ones, are another manifestation of the constriction and control indicated by their low "Anxiety", high "Repression" scores on the MMPI. In contrast, the preferences of the Hi-Lo group were more like those of artists. Finally, the Hi-Hi group showed a dislike for freehand, messy, and complex figures, which Pepper also sees as consistent with their scores on the MMPI A and R scales.

Another study employing both the WFPT and the MMPI was conducted by Wahba (1962). Two-hundred undergraduate males were classified into four groups of 50 students each according to their scores on the A and R scales of the MMPI. Group I was high on A and low on R, Group II was high on A and high on R, Group III was low on A and low on R, and Group IV was low on A and high on R. A comparison of their responses to the 400 items on the WFPT indicated that Groups I and II combined (high A) preferred shaded and freehand, complex designs and rejected ruled, simple designs. These findings are consistent with those of Pepper (1957). On the other hand, Groups II and IV combined (high R) tended to dislike items having male, combined or mixed sex symbols.

In another analysis of his data, Wahba (1962) performed one-way and two-way multivariate analyses of variance on the total number of WFPT items, confirming the significance of the discriminant set of scales and supporting the hypothesis that the A and R scales of the MMPI are orthogonal in their effects. A further investigation with a more homogeneous sample of 95 males corroborated the findings of significant relationships between the WFPT scales and the MMPI R Scale obtained in the first study. The majority of the correlations of WFPT and the MMPI A dimension, however, did not hold up under cross-validation.

Lim and Ullman (1961) used chi-square tests to analyze the relationships among the scores of 125 male neuropsychiatric patients on four WFPT scales and 12 MMPI scales, with the following results. The relationships of scores on the Don't Like (DL), Conformance (CF), Revised Art (RA), and Neuropsychiatric (NP) scales of the WFPT to age, education, and IQ were all negligible. However, the chi-squares relating CF to DL ($\chi^2 = 13.97$), CF to RA ($\chi^2 = -24.18$), CF to NP ($\chi^2 = 8.19$), and RA to NP ($\chi^2 = -52.51$) were highly significant. The writers indicated that the major result was the finding of a significant association between RA and NP ($r = -.792$, $N = 125$). This high association is interpreted as reflecting the fact that "In mental illness there is a primitivization of the perceptual processes which seems to be tapped by the WFPT" (Lim & Ullman, 1961, p. 6).

Also of interest in the Lim and Ullman (1961) study were the relationships of the MMPI clinical scales to the

WFPT. First, the MMPI Pa Scale and WFPT CF Scale were negatively associated (χ^2 = -4.84, N = 124). In addition, the MMPI F Scale was significantly related to the RA (χ^2 = 3.56) and NP (χ^2 = -24.78) scales of the WFPT. Finally, the WFPT RA Scale was moderately related to the MMPI F (χ^2 = 3.56), D (χ^2 = 2.65), and Mf (χ^2 = 2.60) scales, but the WFPT DL Scale was not significantly related to any of the MMPI scales. The writers point to the consistency of the finding of a significant positive association between CF and NP with the finding of a negative association between NP and the MMPI F Scale.

WFPT Scores and the California Psychological Inventory

Harris (1962) computed the correlations of the scores of 91 ninth-grade girls, 111 ninth-grade boys, 86 tenth-grade girls, and 97 tenth-grade boys on 17 scales of the WFPT with their scores on the 18 scales of the California Psychological Inventory (CPI), four Trait Rating Scales and Otis IQs. In general, the correlations of the WFPT scales with teachers' ratings of the students on the traits of intellectual efficiency, conformity, emotional stability, and social relationships were smaller than the correlations of the CPI scales with the trait ratings. In addition, the correlations between the 17 WFPT scales and the 18 CPI scales were rather small. There was a small, but significant, negative correlation between the WFPT RA Scale and the CPI Achievement via Conformance Scale, but RA Scale scores were not significantly related to IQ. These findings led Harris to question the use of the WFPT by high school guidance personnel in the counseling of adolescents. In his opinion, "It appears that personality assessment instruments with more adequately demonstrated reliability and validity than is available so far for the Welsh Figure Preference Test would be more appropriate for use by high school guidance personnel" (Harris, 1961, p. 74).

The Deviation Hypothesis and the WFPT

The "deviation hypothesis", as formulated by Berg (1957), states that: "Deviant response patterns tend to be general; hence those deviant behavior patterns which are significant for abnormality (atypicalness) and thus are regarded as symptoms (earmarks or signs) are associated with other deviant response patterns which are in non-critical areas of behavior and which are not regarded as symptoms of personality aberration (nor as symptoms, signs, earmarks)" (p. 159). In order to test Berg's hypothesis of the generality of deviant response tendencies, Secrest and Jackson (1962) administered a forced-choice version of the WFPT, Berg's Perceptual Reaction Test (PRT), and an Independence-Conformity Inventory to 64 female and 59 male college students. The scoring procedure was to give positive weights to responses which occurred infrequently on the three tests. An analysis of the intercorrelations

among responses to the three tests gave little support to Berg's deviation hypothesis. Scores on the WFPT-Deviation Scale had low, but significant, correlations with PRT-Deviation scores (r = .19) and with the Independence-Conformity Scale (r = .28).

A contention related to Berg's deviation hypothesis is that test item content is irrelevant in personality measurement. In a study designed to test this hypothesis, Norman (1963) compared responses given by groups of male college students to content-irrelevant stimuli with the responses which they gave to content-relevant test items. The predictor variables used in the study were personality descriptive adjectives (Descriptive Adjective Inventory, or DAI), occupational titles (Occupational Preference Inventory or OPI), and figural drawings (WFPT), and the criterion variables were peer ratings on extroversion, agreeableness, dependability, emotional stability, and culture. The WFPT was used as the content-irrelevant task, while the DAI and OPI were viewed as content-relevant. Five empirical scoring keys were devised for each of the three predictors, and scores on these keys were correlated with the five criterion ratings. The findings were that although all five DAI scales were significantly related to the criteria, with only one exception the OPI scales had nearly zero correlations with the criteria, and all five WFPT scales also correlated essentially zero with the criteria. Norman interprets these findings as failing to support Berg's notion of the irrelevance of test item content.

References

Barron, F. Personality style and perceptual choice. J. Pers., 1952, 20, 385-401.

Barron, F. Complexity-simplicity as a personality dimension. J. abnorm. soc. Psychol., 1953a, 48, 163-172.

Barron, F. Some personality correlates of independence of judgment. J. Pers., 1953b, 21, 287-297.

Barron, F. The disposition toward originality. J. abnorm. soc. Psychol., 1955, 51, 478-485.

Barron, F. Originality in relation to personality and intellect. J. Pers., 1957, 25, 730-742.

Barron, F. Creativity and psychological health. Princeton, N.J.: D. Van Nostrand, 1963a.

Barron, F. Discovering the creative personality. In The behavioral sciences and education. New York: College Entrance Examination Board, 1963b, Pp. 79-85.

Barron, F., & Welsh, G. S. Artistic perception as a possible factor in personality style: Its measurement by a figure preference test. J. Psychol., 1952, 33, 199-203.

Berg, I. A. Deviant responses and deviant people: The formulation of the deviation hypothesis. J. counsel. Psychol., 1957, 4, 154-161.

Bieri, J., Bradburn, Wendy M., & Galinsky, M. D. Sex differences in perceptual behavior. J. Pers., 1958, 26, 1-12.

Brown, G. I. An experiment in the teaching of creativity. Sch. Rev., 1964, 72, 437-450.

Brown, G. I. Second study in the teaching of creativity. Harv. educ. Rev., 1965, 35, 39-54.

Brown, G. I. Operational creativity: A strategy for teacher change. Paper presented at Amer. Educ. Res. Assn., Chicago, Feb., 1966.

Buros, O. K. (Ed.). The sixth mental measurements yearbook. Highland Park, N.J.: The Gryphon Press, 1965.

Caracena, P. F., & King, G. F. Generality of individual differences in complexity. J. clin. Psychol., 1962, 30, 588-600.

Cashdan, S., & Welsh, G. S. Personality correlates of creative potential in talented high school students. J. Pers., 1966, 3, 445-455.

Child, I. L. Personal preferences as an expression of aesthetic sensitivity. J. Pers., 1962, 30, 496-512.

Colman, R. W. Comparisons of three creativity measures. Unpublished master's thesis, Univer. of North Carolina, 1966.

Edwards, Elisabeth S. Comparison of responses of children, normal adults and schizophrenics on the Perceptual Maturity Scale. Unpublished doctoral dissertation, Univer. of Denver, 1961.

Falcone, Patricia J. L. An exploratory study designed to ascertain development of creative ability in selected senior high school home economics students. Unpublished doctoral dissertation, Univer. of Wisconsin, 1962.

Golann, S. E. The creativity motive. Unpublished doctoral dissertation, Univer. of North Carolina, 1961.

Golann, S. E. The creativity motive. J. Pers., 1962, 30, 588-600.

Gough, H. Techniques for identifying the creative research scientist. In The creative person. Berkeley: Institute of Personality Assessment and Research, 1961.

Harris, T. L. An analysis of the responses made by adolescents to the Welsh Figure Preference Test and its implications for guidance purposes. Unpublished doctoral dissertation, Univer. of North Carolina, 1961.

Hesterly, S. O., & Berg, I. A. Deviant responses as indicators of immaturity and schizophrenia. J. consult. Psychol., 1958, 22, 389-393.

James, Grace R. The relationships of teachers' characteristics and pupil creativity. Unpublished doctoral dissertation, Univer. of North Carolina, 1963.

L'Abate, L., Boelling, G. M., Hutton, R. D., & Mathews, D. L. The diagnostic usefulness of four potential tests of brain damage. J. consult. Psychol., 1962, 26, 479. (Abstract)

LaCrosse, E. R. The relationship of biological sex, psychological sex, intelligence, and perceptual rigidity to visual exploratory behavior in second and sixth grade, white, southern rural children. Unpublished doctoral dissertation, Univer. of North Carolina, 1965.

Lessler, K. J. Cultural and freudian dimensions of sexual symbols. J. consult. Psychol., 1964, 28, 46-53.

Lim, D. T. Some psychodynamic factors of Chinese character-formation. Unpublished paper, Available from Dr. D. T. Lim, VA Hospital Palo Alto.

Lim, D. T., & Ullmann, L. P. A contribution to Welsh Figure Preference Test norms. Res. Repts. of VA Palo Alto, No. 14, May 1961.

Littlejohn, Mary T. A comparison of responses of ninth-graders to measures of creativity and masculinity-femininity. Unpublished doctoral dissertation, Univer. of North Carolina, 1966.

MacKinnon, D. W. Creativity in architects. In The creative person. Berkeley: Institute of Personality Assessment and Research, 1961a.

MacKinnon, D. W. Fostering creativity in students of engineering. J. eng. Educ., 1961b, 52, 129-142.

McWhinnie, H. J. Review of some research on aesthetic measure and perceptual choice. Stud. Art Educ., 1965, 6, 34-41.

Martin, B. Intolerance of ambiguity in interpersonal and perceptual behavior. J. Pers., 1954, 22, 494-503.

Millman, Marcia, & Chang, Teresa. Inter-correlations among three widely used art tests. Percept. mot. Skills, 1966, 23, 1002.

Moyles, E. W., Tuddenham, R. D., & Block, J. Simplicity/complexity or symmetry/asymmetry? A reanalysis of the Barron-Welsh Art Scales. Percept. mot. Skills, 1965, 685-690.

Norman, W. T. Relative importance of test item content. J. consult. Psychol., 1963, 27, 166-174.

Pepper, L. J. The relationship between Welsh Figure Preference Test responses and indices of anxiety and repression. Unpublished master's thesis, Univer. of North Carolina, 1957.

Pine, F. Creativity and primary process: Sample variations. J. nerv. ment. Dis., 1962, 134, 506-511.

Raychaudhuri, M. Artistic perception of art (painting) and non-art students: A comparative study. Proc. 48th Ind. Sc. Cong., 1961, Part III, p. 571. (Abstract)

Raychaudhuri, M. Complexity-simplicity dimension and perceptual preference pattern of the musician and non-musician. Proc. 49th Ind. Sc. Cong., 1962, Part III, p. 520. (Abstract)

Raychaudhuri, M. Some perceptual characteristics of incipient artists. Indian J. Psychol., 1963, 38, 13-17.

Raychaudhuri, M. Perceptual preferance pattern and creativity. Indian J. appl. Psychol., 1966a, 3, 67-70.

Raychaudhuri, M. Studies in artistic creativity: Personality structure of the musician. Calcutta: Express Printers Private, Ltd., 1966b.

Raychaudhuri, M. Creativity and personality: A review of psychological researches. Indian Psychol. Rev., 1966c, 2, 91-102.

Rosen, J. C. The Barron-Welsh Art Scale as a predictor of originality and level of ability among artists. J. appl. Psychol., 1955, 39, 366-367.

Rosenberg, B. G., & Zimet, C. N. Authoritarianism and aesthetic choice. J. soc. Psychol., 1957, 46, 293-297.

Schultz, K. V., & Knapp, W. E. Perceptual preferences and self descriptions. Personnel & guid. J., 1959, 37, 581-584.

Sechrest, L., & Jackson, D. N. Social intelligence and accuracy of interpersonal predictions. J. Pers., 1961, 29, 167-182.

Sechrest, L., & Jackson, D. N. The generality of deviant response tendencies. J. consult. Psychol., 1962, 26, 167-171.

Smith, Gail V. Welsh Figure Preference Test performance of preschool children. Unpublished master's thesis, Univer. of North Carolina, 1962.

Van de Castle, R. L. Perceptual immaturity and acquiescence among various developmental levels. J. consult. Psychol., 1962, 26, 167-171.

Van de Castle, R. L. Development and validation of a perceptual maturity scale using figure preferences. J. consult. Psychol., 1965, 29, 314-319.

Wahba, M. A multivariate analysis of a figure preference personality test. Unpublished doctoral dissertation, Univer. of North Carolina, 1962.

Watson, W. G. An analysis of responses to the Welsh Figure Preference Test to evaluate its effectiveness as a measure of mental ability. Unpublished doctoral dissertation, Univer. of North Carolina, 1964.

Welsh, G. S. A projective figure-preference test for diagnosis of psychopathology: I. A preliminary investigation. Unpublished doctoral dissertation, Univer. of Minnesota, 1949.

Welsh, G. S. Preliminary manual, Welsh Figure Preference Test (Research ed.). Palo Alto: Consulting Psychologists Press, 1959.

Welsh, G. S. Comparison of D-48, Terman CMT, and Art Scale scores of gifted adolescents. J. consult. Psychol., 1966, 30, 88.

Wrightsman, L. S., & Cook, S. W. Factor analysis and attitude change. Paper presented at Southeastern Psychological Association convention, Atlanta, April, 1965.

Wrightsman, L. S., Wrightsman, Shirley, & Cook, S. W. Test-retest reliability coefficients for thirty-nine measures of personality, attitude, and response set. Unpublished paper, Nov. 1964. Available from Dr. L. S. Wrightsman, George Peabody College for Teachers.